13

A CLOSER LOOK AT THE WESLEYAN INTERCULTURAL COMPETENCE SCALE

A New Tool for Measuring Study Abroad Outcomes

Steven E. Stemler and Carolyn K. Sorkin

In this case study, we describe the development and use of the Wesleyan Intercultural Competence Scale (WICS), a new instrument for assessing the effectiveness of study abroad programs. We begin by discussing the institutional and political context in which this instrument was developed. We then discuss different approaches to assessing intercultural competence found in the literature. Finally, we summarize the methodology used in our work, as well as some results from early studies with the WICS. Along the way, we discuss false starts and challenges we encountered. We end by reflecting on lessons learned and providing recommendations for future research.

Context

Wesleyan University is a small, private, highly selective liberal arts university located in the state of Connecticut in a small town equidistant from Boston and New York. Wesleyan enrolls 2,900 undergraduate and 200 graduate students, plus 200 students in a Graduate Liberal Studies program. The student body is 52% women and hails from 49 states, the District of Columbia, the Virgin Islands, Puerto Rico, and 52 countries outside the United States.

Thirty-one percent of
Black/African Americ
0.07% Native Americ
6% multiracial); anot
bers (48% women, 17
take scholarly research

Wesleyan has lon
Some 45% of student
leyan does not track s
percentage would rise
globe, with less empha
Three academic depar

As is true at othe
of International Stud
abroad outcomes. Wh
Although Wesleyan re
OIS research showed
campus. Furthermore
experience, namely, n
wonder how and how
cultures and what the

Background

Stemler (2012) used
through academic wo
outcomes" (www.wes
cally reviewing simila
ties and the top 125 l
World Report college r
was the most highly
sities (85% mention
quantitative reasonin
Furthermore, intercu
selective liberal arts u

One of the key ch
consensus regarding t
scholars vary greatly
berg & Changnon, 2
in a wide variety of w
ity." A review of the

13

K AT THE
ULTURAL
CE SCALE

oad Outcomes

nd Carolyn K. Sorkin

nd use of the Wesleyan
nstrument for assessing
begin by discussing the
rument was developed.
tercultural competence
ethodology used in our
1 the WICS. Along the
ered. We end by reflect-
ns for future research.

ve liberal arts university
quidistant from Boston
e and 200 graduate stu-
s program. The student
District of Columbia,
tside the United States.

Thirty-one percent of the classes of 2014–2017 are students of color (7% Black/African American, 8% Asian/Asian American, 10% Latino/Hispanic, 0.07% Native American, 0.1% Native Hawaiian or other Pacific Islander, 6% multiracial); another 8% hold foreign passports. All 385 faculty members (48% women, 17% persons of color) teach undergraduates and undertake scholarly research.

Wesleyan has long enjoyed high participation rates in study abroad. Some 45% of students study abroad for a semester or year. Although Wesleyan does not track summer study abroad, anecdotal evidence suggests the percentage would rise markedly were it to do so. Students fan out across the globe, with less emphasis on Western Europe than at many U.S. institutions. Three academic departments require study abroad for their majors.

As is true at other institutions, five or six years ago Wesleyan's Office of International Studies (OIS) began to recognize a need to assess study abroad outcomes. What were students gaining from their experiences abroad? Although Wesleyan requires students to earn grades for courses taken abroad, OIS research showed that these were largely the same as grades earned on campus. Furthermore, grades reflect only a small part of the study abroad experience, namely, mastery of content inside the classroom. We began to wonder how and how well students were immersing themselves in their host cultures and what they were taking away from the experience.

Background

Stemler (2012) used Wesleyan's 10 "Essential Capabilities," to be developed through academic work at Wesleyan and sometimes referred to as "essential outcomes" (www.wesleyan.edu/capabilities), as a framework for systematically reviewing similar statements from the top 125 U.S. national universities and the top 125 liberal arts colleges identified by the 2010 *U.S. News & World Report* college rankings. The development of intercultural competence was the most highly rated priority among highly selective national universities (85% mentioned it), beating out all other skills, including writing, quantitative reasoning skills, civic engagement, and information literacy. Furthermore, intercultural competence was second to only writing at highly selective liberal arts universities (68%).

One of the key challenges in the study abroad field is that, despite relative consensus regarding the importance of developing intercultural competence, scholars vary greatly in how they define or even name the construct (Spitzberg & Changnon, 2009). Indeed, the construct of "culture" itself is defined in a wide variety of ways that do not necessarily equate simply to "nationality." A review of the literature reveals that the variety of existing approaches

to assess intercultural competence can be broadly categorized into measures of knowledge, attitudes, and behaviors.

Assessing Intercultural Competence as a Knowledge-Based Construct

Some researchers, such as Hirsch (1987), have conceptualized intercultural competence as the mastery of a particular body of knowledge. To operationalize this type of approach, Corbitt (1998) developed the Global Awareness Profile (GAP) test, which consists of 120 multiple-choice items designed to test cultural knowledge in 13 areas. These include one general section, six sections related to geographic knowledge of different regions (e.g., Asia, Africa, North America), and six separate sections dealing with knowledge related to the broad contexts of (a) environment, (b) politics, (c) geography, (d) religion, (e) socioeconomics, and (f) culture.

The knowledge-based approach to assessing intercultural competence has several limitations within the context of study abroad assessment. Chief among these is that knowledge-based assessments are necessarily domain specific. In other words, the knowledge one acquires by visiting Japan relates primarily to Japan. What we were seeking instead was an assessment of domain-general skills, those we would expect anyone studying abroad in any culture to acquire and that tend to focus on basic cognitive process elements (e.g., recognition of patterns, recognizing when one has made a faux pas or said something offensive, understanding how to adapt one's habits to thrive in another context).

In addition, within the context of higher education, perhaps the least important elements acquired from being exposed to culturally diverse friends are the benign details of etiquette (e.g., which cheek to kiss when greeting someone, how close to stand to someone, etc.). To be certain, there is a place for this knowledge, but such knowledge does not constitute what universities mean when they speak of a desire to develop students' "intercultural competence." Therefore, the GAP may act as a useful supplement to study abroad assessment but cannot be used alone.

Assessing Intercultural Competence via Attitudes and Behaviors

Two other approaches to measuring intercultural competence involve the assessment of attitudes and specific behaviors. An instrument that successfully integrates both dimensions is the Attitudinal and Behavioral Openness Scale (ABOS) (Caligiuri, Jacobs, & Farr, 2000). The ABOS is based on the theory that personality characteristics, which are stable, enduring traits, will be the best predictors of successful adaptation in a cross-cultural environment. The ABOS consists of 24 items divided across four subscales designed to measure (a) attitudes (e.g., "other cultures fascinate me"), (b) past experience (e.g.,

"I am fluent in another language"), (c) comfort with differences (e.g., "my friends' ethnic backgrounds are different than mine"), and (d) participation in cultural activities (e.g., "I eat at a variety of ethnic restaurants").

However, this trait conceptualization leaves little room for the development of cognitive skills or personality characteristics that lead to enhanced intercultural competence. With personality traits assumed to be relatively stable, little change over time would be expected.

The Development of the WICS and the Nature of Our Target Audience

Within the study abroad community, we found a tremendous appetite for a practical, economical way to assess study abroad outcomes. Increasingly, study abroad offices were being asked by their institutions to show proof that study abroad provided the outcomes that, for many years, had been claimed without much concrete evidence. What we needed was something domain-general in terms of skills, without a focus on a particular culture. It needed to capture learning by registering changes in perspective rather than simply in attitudes.

One of the most widely used attitudinal tests of intercultural competence, based on Bennett's (1986) developmental model of intercultural sensitivity (DMIS), is the Intercultural Development Inventory (IDI) (Hammer, 1999; Hammer, Bennett, & Wiseman, 2003). The IDI consists of 60 general statements to which participants rate their agreement or disagreement on a 7-point scale. Examples of items include "People should avoid individuals from other cultures who behave differently" and "Cultural differences are less important than the fact that people have the same needs, interests, and goals in life." The main advantage of the IDI is that it has a strong theoretical basis that assumes intercultural competence can be developed over time and experience. It has been used to measure the development of intercultural competence within the context of study abroad (Engle & Engle, 2004), and it has been shown to have reasonably good psychometric properties (Paige, Jacobs-Cassuto, Yershova, & DeJaeghere, 2003). The IDI suffers from certain limitations, however. Specifically, it is a self-report measure, which increases its susceptibility to faking (van de Mortel, 2008). It asks abstract attitudinal questions rather than relate each question to a specific context—an approach to questioning that has been shown to relate poorly to actual behavior (DeVellis, 1991). Finally, the IDI is a commercial test and is costly for institutions of higher education to administer and score.

Stemler had been working with a different approach to assessment, known as *situational judgment tests*, for quite some time (Stemler, Elliott, Grigorenko, & Sternberg, 2006; Stemler & Sternberg, 2006) and saw this

method as a way to preserve the theoretical advantages of Bennett's model while overcoming some of the technical limitations of the IDI. Situational judgment tests are an approach to assessment in which individuals are presented with a specific scenario and asked to rate the extent to which they endorse several specific alternatives for responding to the scenario.

Goals

When one is developing a new instrument, the first step is to establish its reliability and validity. After that, one can proceed to use it to answer substantive research questions. Therefore, we had two initial goals for our study. First, we wanted to examine the psychometric characteristics of the WICS (reliability and validity) for use in the assessment of study abroad. Second, while we were interested in a number of substantive research questions, we primarily wanted to know if WICS could measure change in the development of intercultural competence over time.

From there we had all sorts of questions. Are there gender differences from the get-go and in terms of gains? Do people who start lower on the intercultural development scale end up gaining more than those who start higher? Do they gain at the same rate? Do those who start higher gain more? Are there differences among the types of study abroad programs (island, hybrid, direct enrollment)? If a student speaks the host country's language, and this is a different language than his or her native tongue, does this relate to the development of intercultural competence? Next, we briefly summarize the findings from our initial efforts at instrument validation and our study designed to measure change over time.

Methods

The WICS was developed to build on Bennett's theoretical work but overcome the limitations associated with the IDI.

The WICS presents 16 situations that study abroad students are likely to encounter (e.g., making a trip to the grocery store, navigating local transportation; see Stemler, Imada, & Sorkin, 2014, for the list of situations and the frequency with which they were encountered). Associated with each situation are six different response options designed to reflect the six levels of intercultural competence proposed by Bennett (1986). Participants were asked to rate each response on a scale of 1 (*very inaccurate*) to 5 (*very accurate*) to indicate the degree to which each response statement described their actual responses during their most recent experience abroad. Participants

who had not encountered the situations chose "did not encounter such a situation" and did not rate the responses. Table 13.1 provides an example of how the items were mapped onto the DMIS (see Stemler et al., 2014, for the complete instrument and scoring key).

The WICS score is computed by weighing the developmental stage scores differently using the following formula: WICS Score = (–2.5 × Denial) + (–1.5 × Defense) + (–0.5 × Minimization) + (0.5 × Acceptance) + (1.5 × Adaptation) + (2.5 × Integration) + 18. The weights are useful for detecting the relative importance an individual places on the responses associated with each stage. In other words, the weights capture the relative positioning of a given individual and account for the fact that different individuals may use the scale values differently. What really matters is whether an individual states that a particular item in a situation is more or less characteristic of his or her own behavior (e.g., with this approach to scoring, a rating of 4 on an item, reflecting the *adaptation* stage, a rating of 2 on another item, reflecting the *denial* stage, and ratings of 3s on all subsequent items would suggest that the individual is actually closer to the adaptation stage than the acceptance stage). Ultimately, an individual score can be computed for each situation based on this formula. The scores across each of the 16 situations are then added together and averaged. Theoretically, the possible scores for the scale range from –18 to 18, with higher scores indicating greater intercultural competence. A constant value of 18 is added to all scores to eliminate the possibility of negative WICS scores and to make scale interpretation more user-friendly. Thus, the potential scores range from 0 to 36. Scores indicate DMIS categories as follows: denial (0–6), defense (7–12), minimization (13–18), acceptance (19–24), adaptation (25–30), and integration (31–36). See Stemler et al. (2014) for a worked scoring example.

Assessment Findings

Content-Related Validity Evidence

To evaluate the content validity of the test, five undergraduate research assistants not directly involved with the project were recruited as coders. They were provided with a brief description of the six stages of intercultural competence development proposed by Bennett (1986) and asked to independently categorize each response for each situation into one of the six theoretical stages (denial, defense, etc.). The research assistants were deliberately chosen as coders because, unlike experts, they had no prior knowledge of Bennett's DMIS or similar constructs. Thus, their judgment would be based solely on the information they received. This process, which would make the testing of

TABLE 13.1

Example WICS Items Mapped Onto Bennett's Developmental Model of Intercultural Sensitivity

Bennett's Stage	*Description of the Stage*	*Response Example*
Denial	Individuals deny the existence of other cultures or the difference between them.	I tried to go shopping with other Americans or find a store that catered to Americans.
Defense	Individuals react against the threat of other cultures by denigrating the other cultures and promoting the superiority of one's own culture.	I just got the things that were usually sold in the United States because I was afraid of wasting money for something terrible.
Minimization	Individuals acknowledge cultural differences on the surface but consider all cultures as fundamentally similar.	I found that the stores were pretty much like the ones in the United States, and I did not find any big differences. Grocery stores are just grocery stores anywhere.
Acceptance	Individuals accept and respect cultural differences with regard to behavior and values.	I enjoyed finding things that I never saw in the United States, and I was curious about what they were.
Adaptation	Individuals develop the ability to shift their frame of reference to other culturally diverse worldviews through empathy and pluralism.	I often bought local products that were a little different from the products I get in the United States, and I used them a lot in my everyday living.
Integration	Individuals expand and incorporate other worldviews into their own worldview.	I bought and tried local products and discovered really good ones. So I became more open-minded and less restricted by familiarity and brand names when choosing right products for myself.

Note. Adapted from "A Developmental Approach to Training for Intercultural Sensitivity," by M. J. Bennett, 1986, *International Journal of Intercultural Relations, 10(2)*, 179–195.

intercoder agreemen
field of content anal
to prevent the buildu
specialized knowledg
ity (Stemler, 2004) a
Kappa = .86), provid

Criterion-Related

The WICS score fc
SD = 2.74). The co
eight validity measur
items often referred t
who were not U.S. n

As expected, the
taking (r = .25, p <
(r = .39, p < .001) a
.01) and smugness (r
and openness, patric
indicated by the hig
study abroad particip
tude of the correlati
therefore prevented
restricted range (Co
relation value becom

Construct-Related

To examine the con
among the six stage s
and defense) showed
tant from one anot
correlations, indicati
the six developmenta

Factors Associated

In addition to the re
lected participants'
abroad, program typ
factors might be asso

Of the 16 situa
situations participar

intercoder agreement more conservative, is recommended by authors in the field of content analysis (Holsti, 1969; Krippendorff, 2003; Stemler, 2001) to prevent the buildup of shared meaning that can occur among experts with specialized knowledge. The average consensus estimate of intercoder reliability (Stemler, 2004) across all raters and items was 77% agreement (Cohen's Kappa = .86), providing strong evidence for the instrument's content validity.

Criterion-Related Validity Evidence

The WICS score for the sample was normally distributed (M = 25.42, SD = 2.74). The correlation between the WICS score and the scores from eight validity measures are shown in table 13.2. Because the validity measure items often referred to the United States as the home country, six participants who were not U.S. nationals were excluded from the analyses.

As expected, the WICS score was positively correlated with perspective-taking (r = .25, p < .05), empathy (r = .27, p < .01), and internationalism (r = .39, p < .001) and negatively correlated with nationalism (r = –.36, p < .01) and smugness (r = –.32, p < .01). Correlations between the WICS score and openness, patriotism, and ambiguity tolerance were not significant. As indicated by the high mean score of openness (M = 4.27), it is possible that study abroad participants are already high in openness, such that the magnitude of the correlation between these two variables was likely restricted and therefore prevented from reaching statistical significance. If a correction for restricted range (Cohen, Cohen, West, & Aiken, 2003) is applied, the correlation value becomes statistically significant (r = .27, p < .01).

Construct-Related Validity Evidence

To examine the construct validity of the scale, we analyzed the correlations among the six stage scores. Overall, the stages closer to one another (e.g., denial and defense) showed larger positive correlations, whereas the stages more distant from one another (e.g., denial and integration) showed larger negative correlations, indicating that the response items represented well the order of the six developmental stages proposed by Bennett's DMIS.

Factors Associated With Intercultural Competence

In addition to the responses on the WICS and validity measures, we also collected participants' demographic data, their language use during their time abroad, program types, and students' prior experience abroad to learn what factors might be associated with individuals' level of intercultural competence.

Of the 16 situations presented in the instrument, the mean number of situations participants experienced was 11.96 (SD = 2.45), suggesting that

	worldviews through empathy and pluralism.	everyday living.
Integration	Individuals expand and incorporate other worldviews into their own worldview.	I bought and tried local products and discovered really good ones. So I became more open-minded and less restricted by familiarity and brand names when choosing right products for myself.

Note. Adapted from "A Developmental Approach to Training for Intercultural Sensitivity," by M. J. Bennett, 1986, *International Journal of Intercultural Relations, 10(2),* 179–195.

TABLE 13.2
Means of Eight Validity Measures and Correlation With WICS Score (U.S. National Participants Only)

	Openness	*Perspective-Taking*	*Empathy*	*Patriotism*	*Internationalism*	*Nationalism*	*Smugness*	*Ambiguity Tolerance*
Mean	4.27	3.80	3.89	3.29	3.92	1.86	1.97	10.89
(*SD*)	(0.54)	(0.60)	(0.65)	(0.63)	(0.69)	(0.73)	(0.78)	(3.43)
Correlation with WICS score	.06	.25*	.27**	−.09	.39***	−.36**	−.32**	.13

*$p < .05$, **$p < .01$, ***$p < .001$.

the situations wer
was not surprising
programs (n = 73
ations (M = 12.4
abroad programs
d = .84). The nu
the 16 on the tes
score (M = 11.96,
the number of sit
taking local transp
intercultural comp

In addition, t
(English-speaking
correlated with th
Finally, female pa
higher intercultur
SD = 2.12, t[93] =
type, and particip
related to differen

Assessment of Ch

Thirty participan
end of their seme
to see if they cap
The length betw
(M = 101.87, SD

As shown in
the study abroad
of the program, i
competence withi
In contrast, most
differences. The c
ference was ambi
table 13.3).

We also teste
type, prior experi
in the WICS scor
haps because of t
so, the length bet
situations experie
with the size of

the situations were commonly encountered by study abroad participants. It was not surprising that those who had completed one-semester study abroad programs (n = 73) reported that they encountered significantly more situations (M = 12.45, SD = 2.24) than those who had just begun their study abroad programs (n = 24, M = 10.46, SD = 2.50, t[95] = 3.67, $p < .001$, d = .84). The number of different situations participants experienced (of the 16 on the test) was significantly positively associated with the WICS score (M = 11.96, SD = 2.45, r = .37, $p < .001$), suggesting that the greater the number of situations students experienced when studying abroad (e.g., taking local transportation, attending local sporting events), the higher the intercultural competence score.

In addition, the percentage of time spent speaking the local language (English-speaking countries were excluded) was also significantly positively correlated with the WICS score (M = 45.9%, SD = 30.5%, r = .37, $p < .001$). Finally, female participants (M = 7.88, SD = 2.82) showed significantly higher intercultural competence than their male counterparts (M = 6.57, SD = 2.12, t[93] = 2.21, $p < .03$, d = 1.86). Differences in ethnicity, program type, and participants' prior study abroad experience were not significantly related to differences in intercultural competence.

Assessment of Changes in Intercultural Competence

Thirty participants responded to the survey twice (at the beginning and the end of their semester abroad). Their WICS scores were therefore examined to see if they captured changes in participants' intercultural competence. The length between the two responses ranged between 45 and 185 days (M = 101.87, SD = 36.16).

As shown in table 13.3, the WICS score of participants at the end of the study abroad program was significantly higher than at the beginning of the program, indicating that the participants increased their intercultural competence within a short period of time. The effect size was large (d = .71). In contrast, most of the eight validity measure scores showed no significant differences. The only measure besides the WICS that showed significant difference was ambiguity tolerance, with a moderate effect size (d = .47, see table 13.3).

We also tested whether any available variables (e.g., gender, program type, prior experience abroad) significantly predicted the size of the change in the WICS score. None of the variables were found to be significant, perhaps because of the small sample size. However, although only marginally so, the length between Time 1 and Time 2 and the change in the number of situations experienced between Time 1 and Time 2 were positively correlated with the size of change in the WICS score (r = .31 and .35, respectively;

TABLE 13.3
Score Changes in WICS Score and Eight Validity Measures ($N = 30$)

	Time 1	*Time 2*	*Score Change*	
	Mean (SD)	*Mean* (SD)	*Mean* (SD)	*Effect Size* (d)
WICS score	25.02 (2.92)	26.85 (2.18)	1.82 (2.11)***	.71
Openness	4.41 (0.28)	4.45 (0.34)	0.05 (0.23)	.15
Perspective-taking	3.78 (0.58)	3.78 (0.56)	0.00 (0.38)	.00
Empathy	3.92 (0.68)	3.99 (0.76)	0.07 (0.47)	.10
Internationalism	3.92 (0.53)	4.11 (0.69)	0.20 (0.60)	.32
Nationalism	1.86 (0.46)	1.89 (0.70)	0.03 (0.58)	.05
Smugness	2.05 (0.54)	2.06 (0.61)	0.01 (0.68)	.02
Patriotism	3.29 (0.46)	3.51 (0.61)	0.23 (0.48)	.42
Ambiguity tolerance	10.45 (3.02)	11.90 (3.21)	1.45 (2.39)*	.47

*$p < .05$, ***$p < .001$.

$p < .10$). That is, intercultural competence increased more for those who stayed longer and experienced more variability of situations abroad than for those whose participation was of shorter duration and/or who experienced lower variability of situations.

Discussion

The data from our initial studies provide strong support for the reliability and the content, criterion, and construct validity of the WICS. The first key substantive finding is that participants who experience a wider variety of situations (i.e., not only using transportation but also interacting more widely and/or deeply with the host community) tend to score higher on the WICS. This is consistent with findings from previous research suggesting that the more exposure students have to a new culture, the more likely they are to learn and develop (Lou & Bosley, 2008; Savicki & Selby, 2008). However, our data suggest that it may not be the amount of time that matters so much to the development of intercultural competence as the variety of experiences a student encounters.

A second substantive finding from this study is that students who reported spending a great amount of time speaking a foreign language while

ABROAD

ty Measures ($N = 30$)

Score Change	
Mean (SD)	*Effect Size* (d)
32 (2.11)***	.71
.05 (0.23)	.15
.00 (0.38)	.00
.07 (0.47)	.10
.20 (0.60)	.32
.03 (0.58)	.05
.01 (0.68)	.02
.23 (0.48)	.42
.45 (2.39)*	.47

ased more for those who
situations abroad than for
and/or who experienced

support for the reliability
f the WICS. The first key
nce a wider variety of situ-
interacting more widely
ore higher on the WICS.
earch suggesting that the
e more likely they are to
z Selby, 2008). However,
ime that matters so much
the variety of experiences

ly is that students who
; a foreign language while

abroad tended to score more highly on the WICS than those who reported spending less time using the language. This suggests that students who embrace the language and immerse themselves in speaking it will be more fully engaged and will develop their intercultural competence at a more rapid rate than those who spend less time speaking the local language. This finding is consistent with much prior research in the area of study abroad (Hoffa, 2007). However, the finding is all the more interesting because the WICS items do not ask about time spent speaking the language. Rather, WICS asks how individuals dealt with relatively more or less common cultural experiences in their chosen country.

In a third substantive finding, women outperformed men on the WICS. This is consistent with other findings in study abroad research (Vande Berg, Connor-Linton, & Paige, 2009) and suggests that there may be gender differences in the development of intercultural competence that warrant further investigation. Specifically, do women also gain at a greater rate than men? Or do they start with more intercultural competence and finish higher, whereas men make greater gains?

Overall, the data from the study suggest that the WICS holds promise as a new tool for measuring the development of intercultural competence during study abroad. The WICS has several advantages, both psychometric and pragmatic, over current measures. First, it is freely available for use by any institution, provided that it is properly cited and acknowledged. Second, it can easily be expanded and adapted, with new situations being added to the scale, provided they are validated. Third, it is easy to administer and score. Fourth, it is sensitive to developmental changes in intercultural competence even over a relatively brief period of time. Fifth, it avoids many of the problems of self-enhancement that plague self-report measures and lead to ceiling effects. Sixth, the WICS can be used to not only evaluate individuals' intercultural competence but also investigate what elements of study abroad programs might play a more important role in fostering participants' intercultural competence. It can also investigate the interaction between participants' personal factors and program characteristics (e.g., a hybrid program is better for advanced language speakers, but an island program is better for beginners). Such information could be used to help existing study abroad programs redesign their content and help students select programs that fit them well. Finally, the WICS may be useful in a professional development context. For example, the instrument may be used to help students reflect on their responses prior to their study abroad departure and debrief upon their return to the home country. The instrument thus can be used to not only select and evaluate programs but also enhance intercultural competence development.

Lessons Learned

The results from our early studies with the WICS are promising; however, they need to be replicated across a wider variety of students. Furthermore, a large-scale study that increases the total number of study abroad participants would allow for the investigation of a broader range of research questions (see section titled "Goals"). Some progress toward this goal has been made by soliciting participation from institutions via electronic mailing lists and professional conferences. Indeed, over the past 1.5 years, we have collected pre- and posttest data from study abroad participants at eight higher education institutions and two major study abroad program providers. We are currently in the process of writing up the results.

The first major lesson learned in our effort to scale up the project was that many study abroad offices are not yet accustomed to using assessment instruments for research purposes. Although most institutions require students to submit a postexperience evaluation as part of their standard educational practices, the perceived logistics of participating in a large-scale research project proved an impediment to many already overworked study abroad offices. In reality, however, because most institutions already assess students as part of their normal educational practices, the project qualifies for "exempt status" and, therefore, for expedited review at most institutions' institutional review boards.

Second, although our preliminary studies provide excellent evidence for the internal validity of the WICS, we remain quite interested in comparing the scores on the WICS to the scores received by the same students on other measures that are currently being used in the study abroad community (e.g., IDI, GAP, ABOS). During the course of our research, we have, at times, encountered resistance from individuals with a financial or professional stake in their own instrument or perspective on assessment. Comparative data would be tremendously useful for further evaluating the validity and utility of the WICS relative to other instruments used for study abroad assessment. We are actively seeking partners interested in participating in such a research project.

A third lesson from our efforts to coordinate such a study is that many study abroad offices do not see the value in participating if they are already paying for and using one of the standard commercial measures. Why ask students to complete a second instrument? We sympathize with this sentiment; however, we would argue that the long-term benefit to their institution and the greater study abroad research community could be substantial. Because both the instrument and the scoring algorithm for the WICS are freely available (see Stemler et al., 2014), evidence that demonstrates that the WICS is providing as much information as commercial tests would suggest that there

is a tremendous cost saving to be found while using an instrument with demonstrated scientific validity.

Finally, perhaps one of the key limitations of the WICS is that the pretest cannot be taken before participants leave for their study abroad sites. The pretest only makes sense once participants have spent one to two weeks in situ.

Conclusion

The thirst for a free instrument that does not contain a proprietary scoring algorithm and that yields specific, useful information usable in the context of training and assessment is tremendous. The WICS meets these criteria, and we encourage its broad use for all those in the study abroad community who feel they, their students, and their institutions might benefit from it. Although Wesleyan was our first point of departure, we are continuing to expand our research across institutions and encourage others to use the WICS and contribute their data to a research repository. We seek additional input and commentary from the study abroad community regarding key questions to address during the next phases of research. It is our hope that a professional organization in study abroad might develop a repository or other mechanism for sharing findings from institutions that use the WICS. Ideally, this would take place via peer-reviewed publications, but even submissions of anonymized data sets to a website would be of great benefit to researchers, including ourselves, wishing to do secondary analyses of these data.

References

Bennett, M. J. (1986). A developmental approach to training for intercultural sensitivity. *International Journal of Intercultural Relations, 10*(2), 179–195.

Caligiuri, P. M., Jacobs, R. R., & Farr, J. L. (2000). The Attitudinal and Behavioral Openness Scale: Scale development and construct validation. *International Journal of Intercultural Relations, 24*, 27–46.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Corbitt, J. N. (1998). *Global Awareness Profile: GAP facilitator's manual.* Yarmouth, ME: Intercultural Press.

DeVellis, R. F. (1991). *Scale development: Theory and applications* (Vol. 26). Newbury Park, CA: Sage.

Engle, L., & Engle, J. (2004). Assessing language acquisition and intercultural sensitivity development in relation to study abroad program design. *Frontiers: The Interdisciplinary Journal of Study Abroad, 10*, 253–276.

Hammer, M. R. (1999). A measure of intercultural sensitivity: The Intercultural Development Inventory. In S. Fowler & M. Fowler (Eds.), *The intercultural sourcebook* (Vol. 2, pp. 61–72). Yarmouth, ME: Intercultural Press.

Hammer, M. R., Bennett, M. J., & Wiseman, R. (2003). Measuring intercultural sensitivity: The Intercultural Development Inventory. *International Journal of Intercultural Relations: Special Training Issue, 27*(4), 421–443.

Hirsch, E. D. (1987). *Cultural literacy: What every American needs to know*. Boston, MA: Houghton Mifflin.

Hoffa, W. W. (2007). *A history of U.S. study abroad: Beginnings to 1965*. Carlisle, PA: Forum on Education Abroad.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.

Krippendorff, K. (2003). *Content analysis: An introduction to its methodology* (2nd ed.). Newbury Park, CA: Sage.

Lou, K., & Bosley, G. (2008). Dynamics of cultural contexts: Meta-level interventions in the study abroad experience. In V. Savicki (Ed.), *Developing intercultural competence and transformation: Theory, research, and applications in international education* (pp. 276–296). Sterling, VA: Stylus.

Paige, R. M., Jacobs-Cassuto, M., Yershova, Y. A., & DeJaeghere, J. (2003). Assessing intercultural sensitivity: An empirical analysis of the Hammer and Bennett Intercultural Development Inventory. *International Journal of Intercultural Relations, 27*, 467–486.

Savicki, V., & Selby, R. (2008). Synthesis and conclusions. In V. Savicki (Ed.), *Developing intercultural competence and transformation: Theory, research, and applications in international education* (pp. 276–296). Sterling, VA: Stylus.

Spitzberg, B. H., & Changnon, G. (2009). Conceptualizing intercultural competence. In D. K. Deardorff (Ed.), *The SAGE handbook of intercultural competence* (pp. 2–52). Thousand Oaks, CA: Sage.

Stemler, S. E. (2001). An overview of content analysis. *Practical Assessment, Research, and Evaluation, 7*(17). Retrieved from http://PAREonline.net/getvn.asp?v=7&n=17

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation, 9*(4). Retrieved from http://PAREonline.net/getvn.asp?v=9&n=4

Stemler, S. E. (2012). What should university admissions tests predict? *Educational Psychologist, 41*(1), 5–17.

Stemler, S. E., Elliott, J. G., Grigorenko, E. G., & Sternberg, R. J. (2006). There's more to teaching than instruction: Seven strategies for dealing with the practical side of teaching. *Educational Studies, 32*(1), 85–102.

Stemler, S. E., Imada, T., & Sorkin, C. (2014). Development and validation of the Wesleyan Intercultural Competence Scale (WICS). *Frontiers: The Interdiscipli-*

nary Journal of Study Abroad, XXIV, 25–47. Retrieved from www.frontiersjournal.com/documents/StemlerImadaSorkinFrontiersXXIV-Fall2014-withappendix.pdf

Stemler, S. E., & Sternberg, R. J. (2006). Using situational judgment tests to measure practical intelligence. In J. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests* (pp. 107–131). Mahwah, NJ: Lawrence Erlbaum.

Van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, 25*(4), 40–48.

Vande Berg, M., Connor-Linton, J., & Paige, M. R. (2009, Fall). The Georgetown Consortium project: Interventions for student learning abroad. *Frontiers: The Interdisciplinary Journal of International Education, 18*, 1–75.